

# An Ontology-based Contextual Pre-filtering Technique for Recommender Systems

Aleksandra Karpus\*, Iacopo Vagliano<sup>†</sup>, Krzysztof Goczyła\*, Maurizio Morisio<sup>†</sup>

\* Faculty of Electronics Telecommunication and Informatics,  
 Gdańsk University of Technology

ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk-Wrzeszcz, Poland

Email: {aleksandra.karpus, krzysztof.goczyła}@eti.pg.gda.pl

<sup>†</sup> Dept. Control and Computer Engineering,  
 Politecnico di Torino,

C.so Duca degli Abruzzi, 24, 10129 Turin, Italy

Email: {iacopo.vagliano, maurizio.morisio}@polito.it

**Abstract**—Context-aware Recommender Systems aim to provide users with the most adequate recommendations for their current situation. However, an exact context obtained from a user could be too specific and may not have enough data for accurate rating prediction. This is known as the data sparsity problem. Moreover, often user preference representation depends on the domain or the specific recommendation approach used. Therefore, a big effort is required to change the method used. In this paper we present a new approach for contextual pre-filtering (i.e. using the current context to select a relevant subset of data). Our approach can be used with existing recommendation algorithms. It is based on two ontologies: Recommender System Context ontology, which represents the context, and Contextual Ontological User Profile ontology, which represents user preferences. We evaluated our approach through an offline study which showed that when used with well-known recommendation algorithms it can significantly improve the accuracy of prediction.

## I. INTRODUCTION

RECOMMENDER Systems are software tools and techniques providing suggestions for items to be of use to a user. Various kind of items can be suggested, such as music tracks, movies and news. Context-Aware Recommender Systems (CARS) are a particular category of recommender systems which exploits contextual information to provide more effective recommendations. For example, in a temporal context, vacation recommendations in winter should be very different from those provided in summer. Or a restaurant recommendation for a Saturday evening with your friends should be different from that suggested for a workday lunch with co-workers [1].

We distinguish three forms of context-aware recommendation process: *contextual pre-filtering*, *contextual post-filtering* and *contextual modeling* [2]. *Pre-filtering* approaches use the current context to select a relevant subset of data on which recommendation algorithm is applied. *Post-filtering* approaches exploit contextual information to select only relevant recommendations returned by some algorithm. *Contextual modeling*

differs from others techniques as it incorporates the context into recommendation algorithm.

Nowadays context information such as time and location are easy to be obtained with modern devices. However, also other parameters may be considered, such as company (*alone, with friends, with girlfriend*) which may be relevant when recommending movies or vacations. In addition, the exact context sometimes can be too narrow, as Adomavicius and Tuzhilin [2] exemplified by considering the context of watching a movie with a girlfriend in a movie theater on Saturday. Using this exact context may be problematic for several reasons. First, certain aspects of the overly specific context may not be significant. For example, user's movie watching preferences with a girlfriend in a theater on Saturday may be exactly the same as on Sunday, but different from Wednesday's. Therefore, it may be more appropriate to use a more general context specification, i.e. weekend instead of Saturday. Second, exact context may not have enough data for accurate rating prediction, which is known as the data sparsity problem. Thus it may be useful to refer to a more general context such as watching a movie with a girlfriend in a movie theater on weekend, watching a movie with someone in a movie theater on weekend and so on.

Additionally, often user preferences and items representation depends on the application domain addressed or on the specific recommendation approach used. Thus, a big effort is required to adapt the recommender system to another domain or to change the approach to use.

In this paper, we address the problems previously mentioned and we focus on the following research questions:

- *Is it possible to represent context by combining different dimensions (such as time, location, mood, etc.) and representing different granularities for each dimension (e.g. the precise time moment, the day of the week or the season)?*
- *Is it possible to represent user preferences and items in such a way that can be adapted to different application domains and combined with different recommendation approaches?*

The second author was supported by a fellowship from TIM.

We present a new approach to represent context and user preferences, which is based on two ontologies: Recommender System Context (RSCtx)<sup>1</sup> which represents the context, and another ontology, Contextual Ontological User Profile (COUP), which represents user preferences. RSCtx is an ontology in a classical sense, while COUP is an ontology build according to Structured-Interpretation Model (SIM) [3] and it consists of multiple ontological modules. Moreover, we propose a new ontology-based contextual pre-filtering method which could be used with existing recommendation algorithms.

We evaluated our approach by means of an offline study with a rating prediction task which showed that the usage of proposed ontologies and pre-filtering technique with recommendation algorithms could significantly improve the accuracy of prediction according to the Mean Absolute Error (MAE) measure.

The rest of the paper is organized as follows: Section II presents related work, Section III introduces our ontology to represent the context, while Section IV addresses the overall recommendation approach and the representation of user preferences. We detail the evaluation process and its results in Section V. Conclusions and future work close the paper.

## II. RELATED WORK

We distinguish related work in works which addressed representation of context and other ontology based recommender systems proposed. The first is presented in Section II-A while the latter is briefly described in Section II-B.

### A. Context Representation

In this section, firstly we address ontology-based context modeling and then we review context representation for recommender systems.

Many context ontologies have been proposed in the context awareness community. There are a number of surveys which review the literature relevant to context modeling in general [4], [5] or focusing on ontology-based models [6], [7]. In addition, Costabello [8] presented and compared a number of ontology-based context models against a set of requirements. This requirements fit also for our purpose therefore in the following we present the requirements and summarize Costabello's comparison, obviously also considering RSCtx.

The relevant context aware and ontology engineering requirements are:

- R1. Domain independence.** A number of context ontologies have been created to model a given domain-specific scenario. Others adopt a domain-independent approach.
- R2. Coverage.** The ontology must guarantee a proper level of completeness for what concerns the desired contextual dimensions. In particular, the model must support multiple context dimensions such as device features, user preferences, location and time.

- R3. Formality.** Some ontology-based context models rely on formal definitions, while others adopt a more intuitive approach.

- R4. Variable Context Granularity.** Certain ontologies model context dimensions at different level of granularity. For example, location might be expressed in terms of latitude and longitude, or with a label assigned to a place (e.g. office, beach, cinema, etc.).

- R5. User Friendliness Evaluation.** Context-aware application developers must spend a reasonable amount of effort dealing with the context model, thus the ontology must be sufficiently easy to adopt and well documented. The presence of a user evaluation campaign to assess such feature is assessed by certain context models.

- R6. Core ontology approach.** The vocabulary must adopt a modular design, thus focusing on modeling core classes and properties that will be extended by third-party domain specialists.

Linked Data<sup>2</sup> is a set of best practice to publish structured data on the Web. The set of data, vocabularies and ontologies which follows these practices made up the Web of Data. Costabello [8] considered also a number of requirements related to the Linked Data principles, which also fit for our purpose:

- R7. Open World Assumption.** The Web of Data is an open environment, and describing context in this scenario must consider third-party extensions unknown beforehand. Extensibility must be obtained with a low effort, thus additions must not impact on the already existing model.

- R8. Lightweight Ontology.** According to Linked Data best practices [9], the goal is to keep ontologies small and simple.

- R9. Reuse of Existing Terms.** Linked Data best practices favor the reuse and the combination of classes and properties of existing vocabularies. This is done to prevent the proliferation of terms and reduce the range of choices when modeling data.

- R10. Availability on the Web.** Web of Data vocabularies are published on the Web, and accessible according to Web of Data best practices. Moreover, they are associated to an HTML page, the "namespace document", whose task is to provide a textual description of the vocabulary rationale, along with classes and properties explanation and examples.

Following these requirements, Costabello [8] compared a number of ontologies which modeled context and proposed PRISSMA<sup>5</sup>, a vocabulary designed to model client generated context data. We present in the following the main feature of this vocabulary and other related works showed in Table I. PRISSMA satisfies the most of the aforementioned requirements, although *variable context granularity* only partially. It miss formality and user friendliness evaluation, but none of the other works satisfies these two. On the contrary, all

<sup>1</sup><http://softeng.polito.it/rsctx/>

<sup>2</sup><http://linkeddata.org>

<sup>5</sup><http://ns.inria.fr/prissma>

TABLE I

A COMPARISON OF ONTOLOGY-BASED CONTEXT MODELS [8]. FULL SUPPORT IS IDENTIFIED BY ●, PARTIAL SUPPORT BY ○, NO SUPPORT BY THE EMPTY CELL.

Work	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
PRISSMA <sup>3</sup> [8]	●	●		○		●	●	●	●	●
DCO <sup>4</sup>	●	○				●	○			●
SOUPA [10]	●	●				●	●		○	
CoOL [11]	●	○		○		●	○			
CONON [12]	●	●		●		●	●			
CoDaMos [13]	●	●				●	●			○
Korpiää et al. [14]	●	○					○			
Hervás and Bravo [15]	●	●				●				
RSCtx	●	●		●		●	●	●	●	●

the works provide coverage and are domain independent, and all but one support (at least partially) the open world assumption. The only other ontology published on the Web is the Delivery Context Ontology (DCO)<sup>6</sup>, a modular and fine-grained vocabulary to model mobile platforms. It does not provide linking with other vocabulary, thus it is not considered a lightweight ontology. The SOUPA ontology [10] is an OWL ontology which is extensible, i.e. support the open world assumption, and reuses external ontologies, but it does not comply with Linked Data principles, for example it is not publicly available on the Web. CoOL [11] is a modular OWL ontology, which is grounded on F-Logic and uses features typically avoided in lightweight ontology. CONON [12] is another modular OWL ontologies, which is not published on the Web and does not reuse existing vocabularies. CoDaMos [13] is an extensible OWL ontology that is available on the Web but no namespace vocabulary is present. It is not lightweight and does not reuse other vocabularies. Korpiää et al. [14] present a context model designed for mobile, context-aware applications. It is general, but does not reuse existing terms and it is not extensible. Hervás and Bravo [15] propose a modular context model composed by independent ontologies which support extensions, although they do not reuse already existing linked data ontologies.

Various works addressed context representation for recommender systems. Abowd et al. [16] distinguish among primary and secondary context: the first can be directly measured, while the second not and needs to be derived from other types of contextual information. Kaminskis and Ricci [17] reviewed literature about contextual music retrieval. They distinguish among environmental, user-related and multimedia context: the first refers to information about the location of the user, the current time, weather, temperature, etc.; the second to information about the activity of the user, the user's demographic information, emotional state; and the third to other types of information the user is exposed to besides music, e.g., text and images. In addition to traditional dimensions (time location etc.) the authors suggested traffic, noise and light level. As multimedia context, they mention text and images. They indicated some cases in which it can be useful consider this kind of context, e.g. for adapting music to text

context as done by Cai et al. [18]. Baltraunas et al. [19] proposed an approach to assess which contextual factors are important and to which degree they influence user ratings. They conducted a study in which users were asked to judge whether a contextual factor influences the rating given a certain contextual condition. In their survey they focus on tourism domain and consider budget, time availability, transport in addition to traditional dimensions. RSCtx supports most of the addressed dimensions and distinguish among user-related and environmental context. It does not address multimedia context, but it considers the device features.

### B. Ontology based Recommender System

It has been proved that ontological user profile improves recommendation accuracy and diversity [20]. More specifically, a number of ontology-based and context-aware recommender system have been proposed. AMAYA allows management of contextual preferences and contextual recommendations [21]. AMAYA also uses an ontology-based content categorization scheme to map user preferences to entities to recommend. News@hand [22] is a hybrid personalized and context aware recommender system, which retrieves news via RSS feed and annotates by using system domain ontologies. User context is represented by a weighted set of classes from the domain ontology. Rodriguez et al. [23] proposed a CARS which recommends Web services. They use a multi-dimensional ontology model to describe Web services, a user context and an application domain. The multi-dimensional ontology model consists of a three independent ontologies: a user context ontology, a Web service ontology and an application domain ontology, which are combined into one ontology by some properties between classes from different ontologies. The recommendation process consists in assigning a weight to the items based on a list of interests in the user ontology. All this works focus on a specific domain and an ad-hoc algorithm, while our approach for representing user preferences is cross-domain and can be applied to different recommendation algorithm.

Hawalah and Fasli [24] suggest that each context dimension should be described by its own taxonomy. Time, date, location and device are considered as default context parameters in the movie domain. It is possible to add other domain specific context variables as long as they have a clear hierarchical

<sup>6</sup><http://www.w3.org/TR/2009/WD-dcontology-20090616/>

representations. Besides context taxonomies, this approaches uses a reference ontology to build contextual personalized ontological profiles. The key feature of this profile is the possibility of assigning user interests in groups, if these interests are directly associated with each other by a direct relation, sharing the same super-class, or sharing the same property.

Other works uses ontologies and taxonomy to improve the quality of recommendations. Middleton et al. [25] uses an ontological user profile to recommend research papers. Both research papers and user profiles are represented through a taxonomy of topics and the recommendation are generated considering topics of interest for the user and papers classified in those topics. Mobasher et al. [26] proposed a measure which combines semantic knowledge about items and user-item rating, while Anand et al. [27] inferred user preferences from rating data using an item ontology. Their approach recommends the items using the ontology and inferred preferences while computing similarities. A more detailed description of ontology based techniques is available in [28] and [29].

### III. THE RECOMMENDER SYSTEM CONTEXT ONTOLOGY

Recommender System Context (RSCtx) extends PRISSMA<sup>7</sup>, a vocabulary based on Dey's definition of context [30]. PRISSMA relies on the W3C Model-Based User Interface Incubator Group proposal<sup>8</sup>, which describes mobile context as an encompassing term, defined as the sum of three different dimensions: user model and preferences, device features, and the environment in which the action is performed. A graph-based representation of PRISSMA is provided Figure 1.

We designed RSCtx following METHONTOLOGY [31], a well know ontology design method. We assumed there is a predefined set of contextual dimensions in a given application, each with a defined set of attributes and we modeled the contextual information relevant to provide recommendations. We did not focus on any particular domain, on the contrary we aimed at reusing the ontology in different applications. As in PRISSMA, the point of view used to describe the context itself is the application point of view, thus we considered, the user itself as part of the context.

We needed a more detailed representation of the environment, in order to consider other contextual dimensions such as the purpose of the user and the weather. Figure 2 shows how `prissma:Environment` has been extended, by adding a number of properties and related concepts. To represent the weather we integrate `hw:WeatherState` from the Weather Ontology<sup>9</sup>. In this ontology the temperature is represented with respect to the room temperature, thus we defined a new class to represent symbolic values of temperature (such as warm, cold, etc.) and an attribute to represent numeric values, as show in Figure 3.

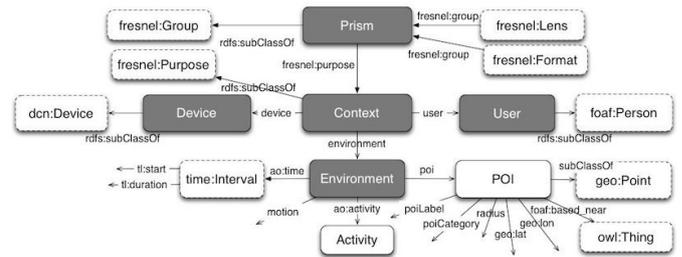


Fig. 1. The PRISSMA vocabulary [8].

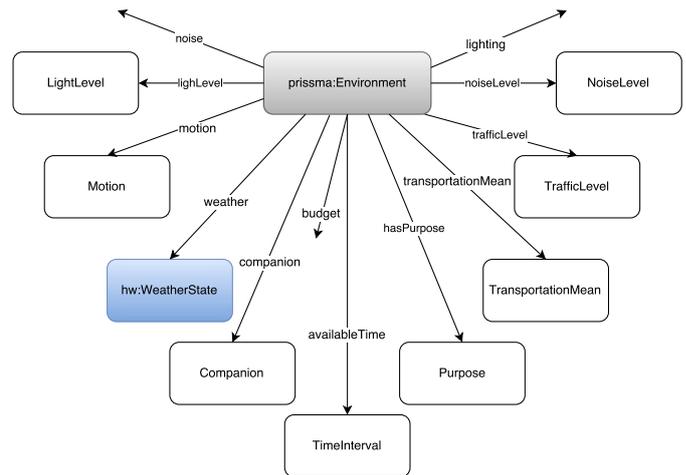


Fig. 2. Relations and concepts which extend `prissma:Environment`

We also extended the time and location representations. We needed a more expressive model of these two dimensions, since asking for recommendations which have the same time stamp and the coordinates of the actual context is too restrictive and the recommender system may not have enough data. On the contrary, by generalizing the context (for example distinguishing among weekend and working day, or considering the city or neighborhood instead of the actual user position) may enable the recommender system to provide recommendations. The concept `prissma:POI` has been extended with various properties to represent the location in the context of a specific site by integrating the Buildings and Rooms vocabulary<sup>10</sup>. Furthermore, other properties related to the hierarchical organization of the location (such as the neighborhood, the city and the province of the current user position) have been added and some concepts from the Juso ontology<sup>11</sup> have been reused. Figure 4 depicts relations and attributes which characterize a location. Yellow rectangles indicate concepts from rooms vocabulary, while blue ones are taken from Juso. The representation of time augments `time:Instant` defined in the Time ontology<sup>12</sup>. Some time intervals have been defined: the hours and the parts of day

<sup>7</sup><http://ns.inria.fr/prissma>

<sup>8</sup><http://www.w3.org/2005/Incubator/model-based-ui/XGR-mbui/>

<sup>9</sup><https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl>

<sup>10</sup><http://vocab.deri.ie/rooms>

<sup>11</sup>[rdfs.co/juso/latest/html](http://rdfs.co/juso/latest/html)

<sup>12</sup><http://www.w3.org/2006/time>

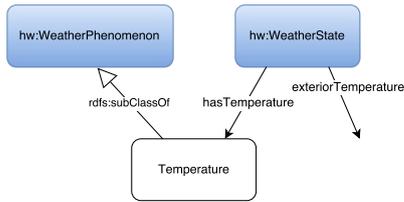


Fig. 3. Temperature representation in our ontology.

(morning, afternoon, etc.). In addition, days of week are classified in weekdays or weekend and seasons are represented. Figure 5 illustrates how time is represented and the relations with PRISSMA and the Time ontology.

Furthermore, we extended the user representation adding some dimensions which may be of interest, as the emotional, mental and physiological state of the user or his fitness. This can be interesting mainly in the medical or fitness domain, but emotional state can affect the user also in taking other kind of decisions, like choosing a movie to watch or music to listen to. Emotional, mental and physiological state concepts are equivalent to emotional, mental and physiological state in the General User Model Ontology (GUMO) [32], an ontology to describe the user which is available on the Web, although it is not compliant with Linked Data principles since it has not a namespace assigned. In addition, the emotional state is an extension of `emoca:Emotion`, which is defined in the Emotion Ontology for Context Awareness (EmOCA)<sup>13</sup>. We added some attributes to the physiological state and also defined an arousal relation which reuse `emoca:Arousal`. Figure 6 depicts the user representation in our ontology.

The emotion in EmOCA are represented according to Russel's model [33]. We extended `emoca:Emotion`, in particular we added pleasure and dominance as subclasses of `emoca:Component` in order to represent emotions according to the Pleasure Arousal Dominance (PAD) model [34] as well, as it is showed in Figure 7. In this way, we can indicate that the emotion is defined by valence and arousal by means of `emoca:isDefinedBy` to refer to Russel's model, while we can indicate that the emotion is defined by pleasure, arousal and dominance to refer to the PAD model. Furthermore, in psychology it is possible to refer to emotion just by indicating its category (such as joy, anger, disgust, etc.). In EmOCA, six categories have been already defined, which can be used also in RSCtx since the emotional state is a subclass of `emoca:Emotion`. We can add more categories in our ontology if it will be needed, although at the moment it has not been done.

#### IV. RECOMMENDATION APPROACH

##### A. Contextual User Profile Ontology

To model user profiles we used the Structured-Interpretation Model (SIM) [35], [36], which consists of two types of ontological modules, i.e. *context types* and *context instances*.

Context types describe the terminological part of an ontology (TBox) and are arranged in hierarchy of inheritance. Context instances describe assertional part of an ontology (ABox) and are connected with corresponding context types through a relation of instantiation. There is another kind of relation, i.e. aggregation, which links context instances of more specific context types to a context instance of a more general context type. In the class hierarchy in a classical ontology there always exists a top concept, i.e. *Thing*. In SIM ontology there is a top context type and a top context instance connected by instantiation. It is possible to add multiple context instances to one context type and aggregate multiple context instances into one context instance. The idea of SIM is shown in Figure 8.

The idea of adaptation SIM ontology as a user profile was proposed by Karpus and Goczyła [37]. They modeled contextual user profiles using only three context variables, i.e. location, time and mood, which influences a split of terminology into ontological modules. Our approach is different in some crucial aspects. First of all, we allow storage of many user profiles in one SIM ontology. We also support a storage of preferences from multiple domains by adding context types related to different domains. Another difference is the number of context variables permitted. We add context types and context instances related to contextual parameters in a dynamic way. As a consequence, we can use as many variables as needed in our approach. An example of contextual profile for one user is shown in Figure 9.

Only three modules in the example illustrated in Figure 9 are fixed: `topContextType`, `topContextInstance` and `UserType`. All others are configurable or can be added in a dynamic way. In `topContextType` we defined the concept `Rating` and its corresponding roles, e.g. `isRatedWith` and `hasValue`. `UserType` is artificial and is present in the SIM ontology because it enable to add many user profiles to the ontology. In the next level of the hierarchy, there are context types that describe domains of interests related to a recommender system which will use the profile. In the next levels, all context types and instances are added to the contextual user profile during the learning phase or later, when a new context situation occurs.

The general process of learning the user profile is as follows. At the beginning there is just the RSCtx ontology and an empty contextual ontology, i.e. with terminological part only. For a given user, an item is taken with the rating and the situation in which it was consumed from the user's history. The level of granularity of the context is checked with the RSCtx ontology and is changed if needed, e.g. shifting from *time = 2 p.m.* to *time = afternoon*. A context instance is created for this context, if it is not already available. Finally, an item with its rating is added to the identified context instance. Each item is represented as a set of individuals of appropriate concepts defined in a domain context type.

##### B. Recommendation

We use the ontologies previously presented for pre-filtering in the recommendation process. The aim is to provide a

<sup>13</sup><http://ns.inria.fr/emoca/>

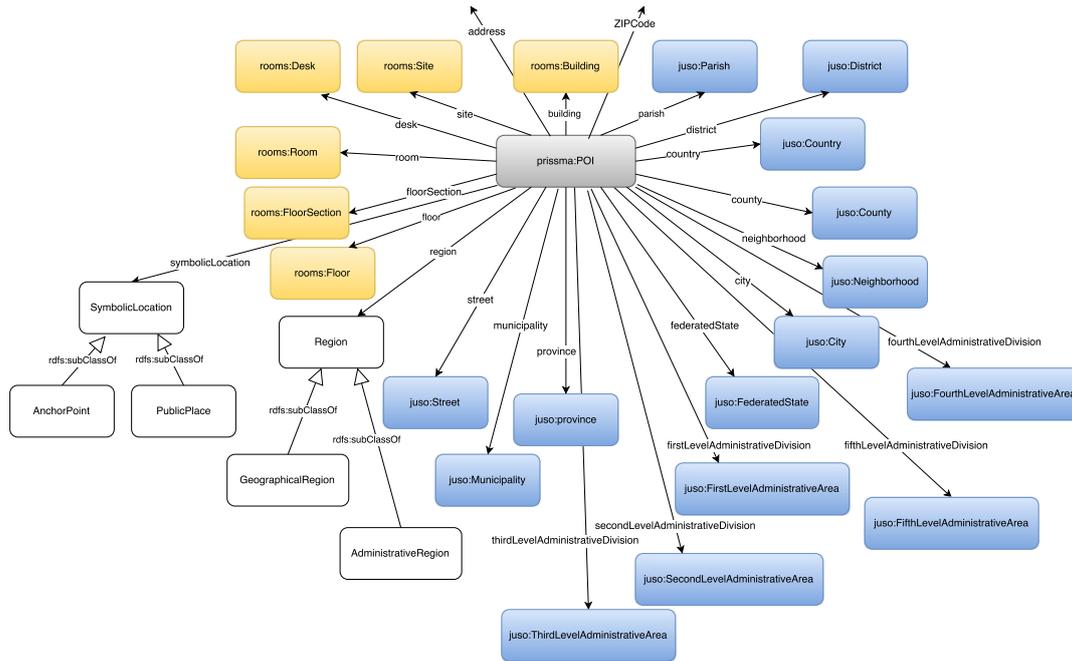


Fig. 4. Concepts and relations of RSCtx representing the location dimension.

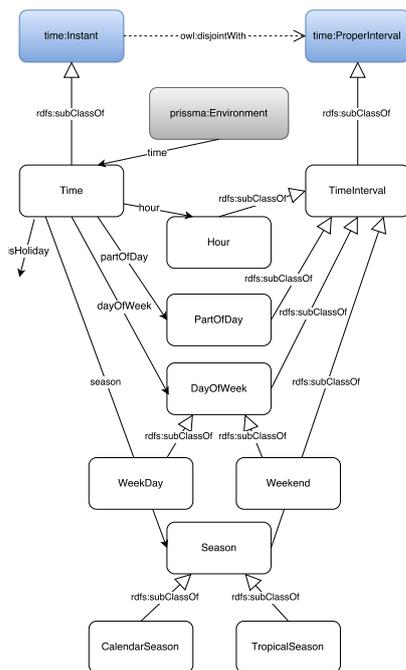


Fig. 5. Time representation in RSCtx ontology as extension of Time and PRISMA ontologies

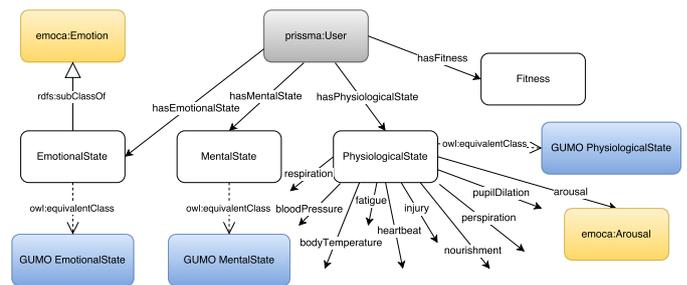


Fig. 6. User representation in RSCtx ontology.

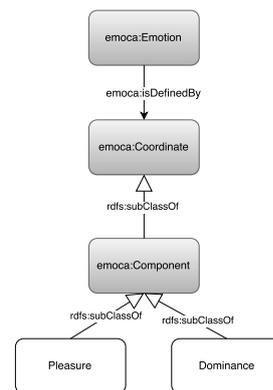


Fig. 7. Emotion representation in RSCtx ontology.

universal context-aware improvement for existing algorithms.

The system consists of three main functional parts: context detection and generalization, user profile and pre-filtering, and recommendation. In the first part, we used the RSCtx ontology to identify the user context from raw data and generalize it in

the desired granularity level. The second part is responsible for building user profile, finding a context instance that fits the actual user context, and returning only relevant user

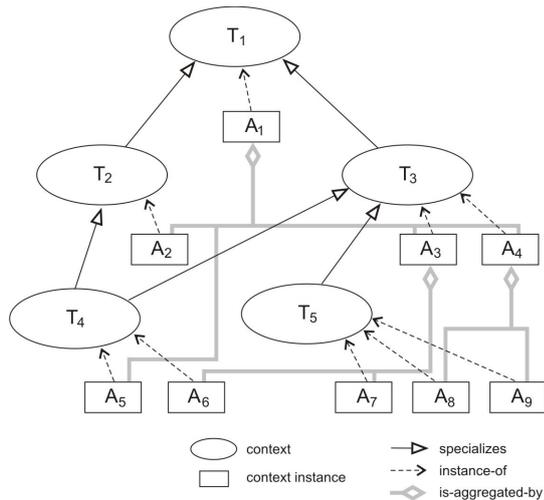


Fig. 8. Structured-Interpretation Model [3]

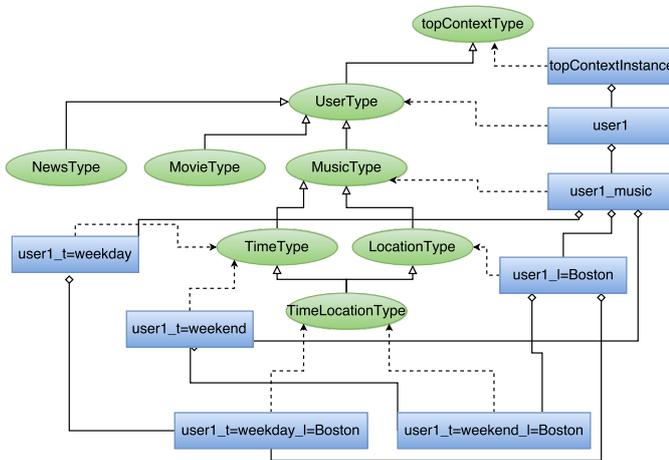


Fig. 9. An example of COUP

preferences. The last part uses well-know algorithms, e.g. Item kNN, User Average, SVD++, for providing recommendations. For this task we exploit implementations from the LibRec<sup>14</sup> library.

The general recommendation process is as follows. Given a user and his current situation, a proper generalization of his context is generated by using the RSCtx ontology. Then, an appropriate context instance from COUP is identified by using the generalized context. If a context instance is not found in the user profile, the generalization step is repeated to search for a module that corresponds to the new context. If it is found, user preferences are prepared to be used with a recommendation algorithm.

## V. EVALUATION

We conducted an offline study in order to evaluate the RSCtx ontology and the COUP ontology. We selected a

number of algorithms and we compared the accuracy of each algorithm when used as is and when combined with proposed ontologies. We aimed to answer the following question: *does our context and user preferences representation improve accuracy of recommendation algorithms?*

We relied on ConcertTweets dataset [38], which combines implicit and explicit user ratings with rich content as well as spatio-temporal contextual dimensions and social network data. It contains ratings that refer to musical shows and concerts of various artists and bands. Since the dataset was generated automatically, there were some duplicated events, for example the same concert was occurring twice, once with country *United Kingdom* and once *UK*. We fixed this kind of situations. Another problem with the dataset is the usage of two rating scales: one numerical scale with ratings in the range [0.0, 5.0] and one descriptive scale with possible values equal to *yes*, *maybe* and *no*, although *no* never occurred. We decided to split the dataset into two separate sets according to the scale type and we mapped the descriptive values *yes*, *maybe* and *no* with the numerical values 2, 1 and 0. Table II presents some statistics about the data by considering the whole dataset and each of the sets generated when splitting by scale type. We prepared two pairs (one for each scale) of training and test sets for hold-out validation. In the test set we put 20% of the newest ratings of each user. All other ratings were placed in the training set. The split was performed based on rating timestamp value.

Because of the dataset domain, we needed to add to contextual user profile a new *context type*, *MusicType*. For this purpose we reused two existing ontologies related to music, i.e. *musicbrainz*<sup>15</sup> and *music vocabulary*<sup>16</sup>. We used their terminological parts only. Any item can be represented as a pair of individuals of type *mb:Artist* and *m:Musical\_Event* and with their corresponding properties.

Our approach is a pre-filtering approach and can be used with existing recommendation methods. We evaluated the ontologies with five algorithms: Random Guess, Item kNN, User Average, SVD++ and Time SVD++. We compared the results of first four algorithms without pre-filtering and with pre-filtering, while the fifth was executed without pre-filtering only, because it already contains the *time* as a contextual factor [39]. We used it as a baseline for comparing our contextual pre-filtering technique combined with SVD++ algorithm.

We performed an experiment for rating prediction task and measured accuracy with MAE. Results are presented in Table III and Figures 10 and 11. It should be noticed that without pre-filtering, the User Average algorithm outperforms SVD++. This may be due to the way in which users rate musical events: it may be possible that they do not use the whole rating scale but just a part of it, e.g. a user evaluates only those events that he likes (his ratings are always greater than 3.0). As it can be seen in Figure 10 and Table III, when our ontological pre-filtering approach is applied, results on the *numerical scale*

<sup>14</sup><http://www.librec.net/>

<sup>15</sup><https://musicbrainz.org/>

<sup>16</sup><http://www.kanzaki.com/ns/music>

TABLE II  
STATISTICS ON THE DATA CONTAINED IN CONCERTTWEETS DATASET AT THE TIME OF THE EXPERIMENT

	All	Descriptive ratings	Numeric Ratings
Number of users	61803	56519	16479
Number of musical events	116320	110207	21366
Number of pairs artist and musical events	137382	129989	23383
Number of ratings	250000	219967	30033
Maximum number of ratings per user	1423	1419	92
Minimum number of ratings per user	1	1	1
Average number of ratings per user	4.045	3.892	1.823
Maximum number of ratings per item	218	216	38
Minimum number of ratings per item	1	1	1
Average number of ratings per item	2.149	1.996	1.406
Number of users who ranked at least 5 items	13241	11548	962
Number of users who ranked at least 10 items	5369	4639	190
Number of users who ranked at least 50 items	289	244	4
Number of users who ranked at least 100 items	66	54	0
Sparsity	0.999971	0.999970	0.999922

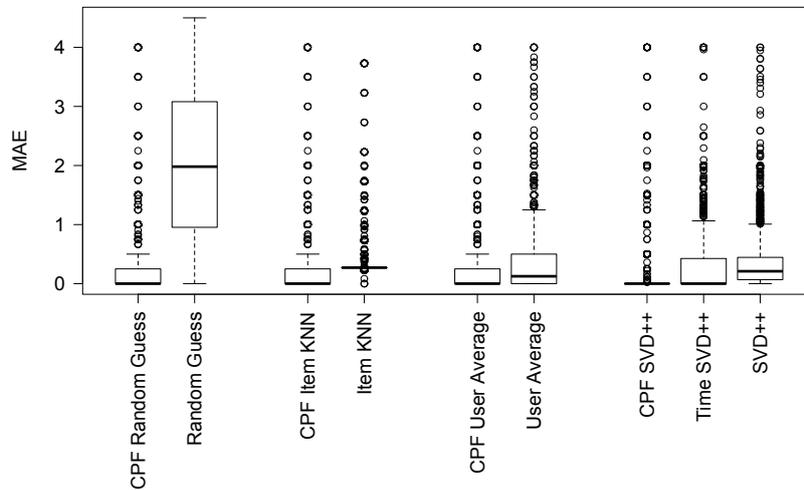


Fig. 10. Boxplots of MAE values of different algorithms computed per user on subset with numeric ratings

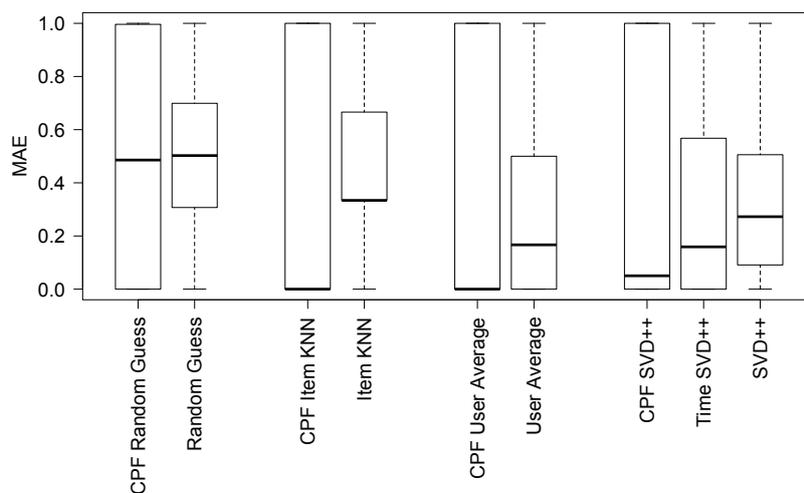


Fig. 11. Boxplots of MAE values of different algorithms computed per user on subset with descriptive ratings

TABLE III  
MAE VALUES COMPUTED FOR WHOLE TEST SETS

Contextual pre-filtering	Numeric ratings		Descriptive Ratings	
	YES	NO	YES	NO
Random Guess	0.2315	2.0998	0.4694	0.4989
User Average	0.2312	0.3026	0.3624	0.2570
Item kNN	0.2312	0.3976	0.3624	0.4374
SVD++	0.2514	0.3511	0.3621	0.3101
Time SVD++	NA	0.2693	NA	0.2975

subset are better. Our contextual pre-filtering combined with classical SVD++ performs better than Time SVD++. There could be two reasons for this behavior. First, the usage of more contextual parameters than just one, the time, gives more improvement to prediction accuracy. Second, our approach (even if used with time parameter only) with SVD++ is truly better than Time SVD++ algorithm. This should be addressed in further work.

From Figure 11 we see that a median value for our approach is improved for all algorithms but the overall MAE value for *descriptive scale* subset is worst for all of the cases. This suggest that in the case of binary scale (*yes/maybe*) contextual pre-filtering may increase the sparsity and noisiness of the data. Thus, the recommendation algorithm may not always predict the rating. However, the difference between results for two subsets could be caused not by wrong pre-filtering method but by psychological differences between *a priori* and *a posteriori* evaluation by a user. It is more reliable when a user evaluates an item after he consumed it than when he declares what he would do or prefer. This could lead us to conclusion that this approach could be successfully applied in recommender systems where numeric scale is used to rate items in *a posteriori* way. Currently, we have not identified any other limitations for applying proposed contextual pre-filtering approach.

To check the statistical significance of the results, we applied Wilcoxon test with *p-value* 0.01. We chose this statistical test because we cannot guarantee the normal distribution of obtained results. The test confirmed the statistical significance of our results.

## VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented a new approach for contextual pre-filtering in Recommender Systems. It is based on two ontologies: Recommender System Context (RSCtx), which represents the context, and Contextual Ontological User Profile (COUP), which represents user preferences. RSCtx extends PRISMA and represents different context dimensions on different granularity levels. COUP was modeled according to SIM approach for modularization. Different users' parts of profile are represented in different ontological module. This allows us to: (I) store multiple users in one ontology, (II) clearly distinguishing user preferences from different domains, but keeping all the user preferences together, and (III) split user interests from one domain into "micro profiles" related

to some contextual situation without loosing the possibility to reason on different level of context granularity.

We successfully applied RSCtx for context identification and generalization tasks, showing that it is possible to represent context by combining different dimensions and representing different granularities for each dimension. We used COUP for representing user preferences in different context in the domain of musical events and for obtaining user data relevant to his current context for rating prediction task with baseline algorithms. This offline study showed that the usage of proposed ontologies with recommendation algorithms could significantly improve the accuracy of prediction according to MAE measure. This confirmed part of the second research question, i.e. that it is possible to represent user preferences and items in such a way that can be combined with different recommendation approaches. The next step in our research is proving that we can adapt our user representation for different domains.

As future work, we plan to extend our experiment to ranking task as well as to investigate on the influence of the proposed approach on diversity and novelty of recommendations.

## REFERENCES

- [1] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 1–35. ISBN 978-0-387-85819-7. [Online]. Available: [http://dx.doi.org/10.1007/978-0-387-85820-3\\_1](http://dx.doi.org/10.1007/978-0-387-85820-3_1)
- [2] G. Adomavicius and A. Tuzhilin, *Recommender Systems Handbook*. Boston, MA: Springer US, 2011, ch. Context-Aware Recommender Systems, pp. 217–253. ISBN 978-0-387-85820-3. [Online]. Available: [http://dx.doi.org/10.1007/978-0-387-85820-3\\_7](http://dx.doi.org/10.1007/978-0-387-85820-3_7)
- [3] K. Goczyla, W. Waloszek, and A. Waloszek, "Contextualization of a DL knowledge base," in *Proc. of the 2007 Int. Workshop on Description Logics (DL2007)*, 2007. [Online]. Available: [http://ceur-ws.org/Vol-250/paper\\_55.pdf](http://ceur-ws.org/Vol-250/paper_55.pdf)
- [4] C. Bettini, O. Brdiczka, K. Henriksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni, "A survey of context modelling and reasoning techniques," *Pervasive and Mobile Computing*, vol. 6, no. 2, pp. 161 – 180, 2010. doi: 10.1016/j.pmcj.2009.06.002 Context Modelling, Reasoning and Management. [Online]. Available: <http://dx.doi.org/10.1016/j.pmcj.2009.06.002>
- [5] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, and L. Tanca, "A data-oriented survey of context models," *SIGMOD Rec.*, vol. 36, no. 4, pp. 19–26, Dec. 2007. doi: 10.1145/1361348.1361353. [Online]. Available: <http://dx.doi.org/10.1145/1361348.1361353>
- [6] R. Kruppenacher and T. Strang, "Ontology-based context modeling," in *In Workshop on Context-Aware Proactive Systems*, 2007.
- [7] J. Ye, L. Coyle, S. Dobson, and P. Nixon, "Ontology-based models in pervasive computing systems," *Knowl. Eng. Rev.*, vol. 22, no. 4, pp. 315–347, Dec. 2007. doi: 10.1017/S0269888907001208. [Online]. Available: <http://dx.doi.org/10.1017/S0269888907001208>
- [8] L. Costabello, *Context-Aware Access Control and Presentation for Linked Data*, 2013, ch. A Declarative Model for Mobile Context, pp. 21–32.
- [9] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, 1st ed. Morgan & Claypool, 2011. ISBN 9781608454303. [Online]. Available: <http://dx.doi.org/10.2200/S00334ED1V01Y201102WBE001>
- [10] H. Chen, T. Finin, and A. Joshi, *Ontologies for Agents: Theory and Experiences*. Basel: Birkhäuser Basel, 2005, ch. The SOUPA Ontology for Pervasive Computing, pp. 233–258. ISBN 978-3-7643-7361-0. [Online]. Available: [http://dx.doi.org/10.1007/3-7643-7361-X\\_10](http://dx.doi.org/10.1007/3-7643-7361-X_10)
- [11] T. Strang, C. Linnhoff-Popien, and K. Frank, *Distributed Applications and Interoperable Systems: 4th IFIP WG6.1 Int. Conf., DAIS 2003, Paris, France, November 17-21, 2003. Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, ch. CoOL: A

- Context Ontology Language to Enable Contextual Interoperability, pp. 236–247. ISBN 978-3-540-40010-3. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-40010-3\\_21](http://dx.doi.org/10.1007/978-3-540-40010-3_21)
- [12] X. H. Wang, D. Q. Zhang, T. Gu, and H. K. Pung, "Ontology based context modeling and reasoning using owl," in *Proc. of the Second IEEE Annual Conf. on Pervasive Computing and Communications Workshops*, ser. PERCOMW '04. Washington, DC, USA: IEEE Computer Society, 2004. doi: 10.1109/PERCOMW.2004.1276898. ISBN 0-7695-2106-1 pp. 18–. [Online]. Available: <http://dx.doi.org/10.1109/PERCOMW.2004.1276898>
- [13] D. Preuveneers, J. Bergh, D. Wagelaar, A. Georges, P. Rigole, T. Clerckx, Y. Berbers, K. Coninx, V. Jonckers, and K. Bosschere, *Ambient Intelligence: Second European Symposium, EUSAI 2004, Eindhoven, The Netherlands, November 8-11, 2004. Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, ch. Towards an Extensible Context Ontology for Ambient Intelligence, pp. 148–159. ISBN 978-3-540-30473-9. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-30473-9\\_15](http://dx.doi.org/10.1007/978-3-540-30473-9_15)
- [14] P. Korpipää and J. Mäntyjärvi, *Modeling and Using Context: 4th Int. and Interdisciplinary Conf. CONTEXT 2003 Stanford, CA, USA, June 23–25, 2003 Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, ch. An Ontology for Mobile Device Sensor-Based Context Awareness, pp. 451–458. ISBN 978-3-540-44958-4. [Online]. Available: [http://dx.doi.org/10.1007/3-540-44958-2\\_37](http://dx.doi.org/10.1007/3-540-44958-2_37)
- [15] R. Hervás and J. Bravo, "Towards the ubiquitous visualization: Adaptive user-interfaces based on the semantic web," *Interacting with Computers*, vol. 23, no. 1, pp. 40 – 56, 2011. doi: <http://dx.doi.org/10.1016/j.intcom.2010.08.002>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0953543810000676>
- [16] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggle, "Towards a better understanding of context and context-awareness," in *Proc. of the 1st Int. Symposium on Handheld and Ubiquitous Computing*, ser. HUC '99. London, UK, UK: Springer-Verlag, 1999. ISBN 3-540-66550-1 pp. 304–307. [Online]. Available: [http://dx.doi.org/10.1007/3-540-48157-5\\_29](http://dx.doi.org/10.1007/3-540-48157-5_29)
- [17] M. Kaminskis and F. Ricci, "Contextual music information retrieval and recommendation: State of the art and challenges," *Computer Science Review*, vol. 6, no. 2–3, pp. 89 – 119, 2012. doi: 10.1016/j.cosrev.2012.04.002. [Online]. Available: <http://dx.doi.org/10.1016/j.cosrev.2012.04.002>
- [18] R. Cai, C. Zhang, C. Wang, L. Zhang, and W.-Y. Ma, "Musicsense: Contextual music recommendation using emotional allocation modeling," in *Proc. of the 15th ACM Int. Conf. on Multimedia*, ser. MM '07. New York, NY, USA: ACM, 2007. doi: 10.1145/1291233.1291369. ISBN 978-1-59593-702-5 pp. 553–556. [Online]. Available: <http://dx.doi.org/10.1145/1291233.1291369>
- [19] L. Baltrunas, B. Ludwig, S. Peer, and F. Ricci, "Context relevance assessment and exploitation in mobile recommender systems," *Personal Ubiquitous Comput.*, vol. 16, no. 5, pp. 507–526, Jun. 2012. doi: 10.1007/s00779-011-0417-x. [Online]. Available: <http://dx.doi.org/10.1007/s00779-011-0417-x>
- [20] Z. Su, J. Yan, H. Ling, and H. Chen, "Research on personalized recommendation algorithm based on ontological user interest model," *J. of Computational Information Systems*, vol. 8, no. 1, pp. 169–181, Jan. 2012.
- [21] C. Rack, S. Arbanowski, and S. Steglich, "Context-aware, Ontology-based Recommendations," in *SAINT-W '06: Proc. of the Int. Symposium on Applications on Internet Workshops*. Washington, DC, USA: IEEE Computer Society, 2006. doi: 10.1109/saint-w.2006.13. ISBN 0769525105 pp. 98–104. [Online]. Available: <http://dx.doi.org/10.1109/saint-w.2006.13>
- [22] I. Cantador, A. Bellogín, and P. Castells, "Ontology-based personalised and context-aware recommendations of news items," in *Proc. of the 2008 IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology - Volume 01*, ser. WI-IAT '08. Washington, DC, USA: IEEE Computer Society, 2008. doi: 10.1109/WIAT.2008.204. ISBN 978-0-7695-3496-1 pp. 562–565. [Online]. Available: <http://dx.doi.org/10.1109/WIAT.2008.204>
- [23] J. Rodríguez, M. Bravo, and R. Guzmán, "Multidimensional ontology model to support context-aware systems," 2013. [Online]. Available: <http://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/view/7187>
- [24] A. Hawalah and M. Fasli, "Utilizing contextual ontological user profiles for personalized recommendations," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4777 – 4797, 2014. doi: <http://dx.doi.org/10.1016/j.eswa.2014.01.039>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417414000633>
- [25] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure, "Ontological user profiling in recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 54–88, Jan. 2004. doi: 10.1145/963770.963773. [Online]. Available: <http://doi.acm.org/10.1145/963770.963773>
- [26] B. Mobasher, X. Jin, and Y. Zhou, *Web Mining: From Web to Semantic Web: First European Web Mining Forum, EWMF 2003, Invited and Selected Revised Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, ch. Semantically Enhanced Collaborative Filtering on the Web, pp. 57–76. ISBN 978-3-540-30123-3. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-30123-3\\_4](http://dx.doi.org/10.1007/978-3-540-30123-3_4)
- [27] S. S. Anand, P. Kearney, and M. Shapcott, "Generating semantically enriched user profiles for web personalization," *ACM Trans. Internet Technol.*, vol. 7, no. 4, Oct. 2007. doi: 10.1145/1278366.1278371. [Online]. Available: <http://doi.acm.org/10.1145/1278366.1278371>
- [28] P. Lops, M. de Gemmis, and G. Semeraro, *Recommender Systems Handbook*. Boston, MA: Springer US, 2011, ch. Content-based Recommender Systems: State of the Art and Trends, pp. 73–105. ISBN 978-0-387-85820-3. [Online]. Available: [http://dx.doi.org/10.1007/978-0-387-85820-3\\_3](http://dx.doi.org/10.1007/978-0-387-85820-3_3)
- [29] T. Di Noia and V. C. Ostuni, *Reasoning Web. Web Logic Rules: 11th Int. Summer School 2015, Berlin, Germany, July 31-August 4, 2015, Tutorial Lectures*. Cham: Springer International Publishing, 2015, ch. Recommender Systems and Linked Open Data, pp. 88–113. ISBN 978-3-319-21768-0. [Online]. Available: [http://dx.doi.org/10.1007/978-3-319-21768-0\\_4](http://dx.doi.org/10.1007/978-3-319-21768-0_4)
- [30] A. K. Dey, "Understanding and using context," *Personal Ubiquitous Comput.*, vol. 5, no. 1, pp. 4–7, Jan. 2001. doi: 10.1007/s007790170019. [Online]. Available: <http://dx.doi.org/10.1007/s007790170019>
- [31] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," in *Proc. Symposium on Ontological Engineering of AAAI, 1997*.
- [32] D. Heckmann, T. Schwartz, B. Brandherm, M. Schmitz, and M. Wilamowitz-Moellendorff, *User Modeling 2005: 10th Int. Conf., UM 2005, Edinburgh, Scotland, UK, July 24-29, 2005. Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, ch. Gumo – The General User Model Ontology, pp. 428–432. ISBN 978-3-540-31878-1. [Online]. Available: [http://dx.doi.org/10.1007/11527886\\_58](http://dx.doi.org/10.1007/11527886_58)
- [33] J. A. Russell, "A circumplex model of affect," *J. of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980. doi: 10.1037/h0077714. [Online]. Available: <http://dx.doi.org/10.1037/h0077714>
- [34] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292. doi: 10.1007/BF02686918. [Online]. Available: <http://dx.doi.org/10.1007/BF02686918>
- [35] K. Goczyła, A. Waloszek, W. Waloszek, and T. Zawadzka, *Intelligent Tools for Building a Scientific Information Platform*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, ch. Modularized Knowledge Bases Using Contexts, Conglomerates and a Query Language, pp. 179–201. ISBN 978-3-642-24809-2. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-24809-2\\_11](http://dx.doi.org/10.1007/978-3-642-24809-2_11)
- [36] K. Goczyła, A. Waloszek, and W. Waloszek, "Towards context-semantic knowledge bases," in *Federated Conf. on Computer Science and Information Systems - FedCSIS 2012, Wroclaw, Poland, 9-12 September 2012, Proc.*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012. ISBN 978-83-60810-51-4 pp. 475–482. [Online]. Available: <https://fedcsis.org/proceedings/2012/pliks/388.pdf>
- [37] A. Karpus and K. Goczyła, "A multi-domain hybrid recommender systems based on a dynamic contextual ontological user profile," in *Doctoral Consortium (IC3K 2014)*, 2014. doi: 10.5220/0005174300830087. ISBN Not Available pp. 83–87. [Online]. Available: <http://www.scitepress.org/DigitalLibrary/PublicationsDetail.aspx?ID=6Kh9MDlu7qs=&t=1>
- [38] P. Adamopoulos and A. Tuzhilin, "Estimating the Value of Multi-Dimensional Data Sets in Context-based Recommender Systems," in *8th ACM Conf. on Recommender Systems (RecSys 2014)*, 2014.
- [39] Y. Koren, "Collaborative filtering with temporal dynamics," in *Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009. doi: 10.1145/1557019.1557072. ISBN 978-1-60558-495-9 pp. 447–456. [Online]. Available: <http://doi.acm.org/10.1145/1557019.1557072>